

ABSTRACT

A processor implemented method of identifying the text genre of a machine-readable, untagged text. The processor implemented method begins by generating a cue vector from the text, which represents occurrences in the text of a first set of nonstructural, surface cues, which are easily computable. Afterward, the processor determines whether the text is an instance of a first text genre using the cue vector and a weighting vector associated with the first text genre.

03100189-061898